Trace transform based method for color image domain identification

Igor G.Olaizola, *Member, IEEE,* Marco Quartulli, *Member, IEEE,* Julián Flórez, *Member, IEEE,* and Basilio Sierra, *Member, IEEE*

Abstract—Context categorization is a fundamental pre-requisite for multi-domain multimedia content analysis applications in order to manage contextual information in an efficient manner. In this paper, we introduce a new color image context categorization method (DITEC) based on the trace transform. The problem of dimensionality reduction of the obtained trace transform signal is addressed through statistical descriptors that keep the underlying information. These extracted features offer a highly discriminant behavior for content categorization. The theoretical properties of the method are analyzed and validated experimentally through two different datasets.

Index Terms—Trace Transform, Image domain identification, CBIR, Pattern recognition.

1 Introduction

The importance of context is very well known in Content Based Image Retrieval (CBIR) [1], [2]. Many low-level features, based on shape, texture, color and other local descriptors broadly used in computer vision incur under multi-domain circumstances in a circular interdependence of feature extractors where *a priori* information is needed to parametrize adequate feature extractors. A possible approach to try and reduce this dependency involves the exploitation of global image context characterization for semantic domain inference. This prior information on scene context can represent a valuable asset in computer vision for purposes ranging from regularization to the pre-selection of local primitive feature extractors [3].

Novel semantic approaches that try to overcome the current existing limitation derived from fixed taxonomies and manual annotations, rely on automatic or semiautomatic ingestion processes. These processes minimize the *semantic gap* by introducing *semantic middleware* [4] layers based on a combination of:

- explicit information provided by human made taxonomies.
- relevance feedback data and knowledge extracted from manual annotations.
- implicit information obtained by data mining techniques through training processes.

This is specially relevant for broad-domain data intensive multimedia retrieval activities like the news production in TV broadcasting sector or large-scale earth observation archive navigation and exploitation.

- *I.G. Olaizola, J. Flórez and M. Quartulli are with the Department of Digital TV & Multimedia Services, Vicomtech, Spain, Donostia–San Sebastián 20009.*
  E-mail: {iolaizola,jflorez,mquartulli}@vicomtech.org
- *Basilio Sierra is with the University of the Basque Country*
  E-mail: b.sierra@ehu.es

1.1 Related works

Research contributions related to the approach proposed in this paper are outlined in this section.

Local features have broadly been used for context categorization [5], [6]. SIFT [7] and SURF [8] are among most popular choices in this respect. A two step approach for the efficient use of local features is proposed by several authors like Ravinovich et al. [9] and Choi et al. [10]. Olaizola et al. [11] propose an architecture for hypothesis reinforcement based on an initial analysis of low-level features for context categorization and further hypothesis creation. This architecture can exploit context specific feature extractors to validate or refuse the initial context hypothesis. This stresses the value of global descriptors.

Among different global descriptors like histograms of several local features [12], texture features, self similarity [13], there are some specific algorithms in the literature which have shown a great potential: GIST [14], [15] is probably the one of the most popular ones. Watanabe et al. [16] propose a global descriptor based on the codewords provided by Lempel-Ziv [17], [18] entropy coders, exploiting the relationship between the complexity of an image and the context to which it may belong. The Ridgelet transform [19], [20], [21] has been successfully used as a global feature for image categorization and handwritten character recognition. In typical operational implementations, all these algorithms are typically combined with other global or local features.

The trace transform has been already used for several computer vision applications. Indeed, a method based on this transform has been included in the MPEG-7 [22] standard specification for image fingerprinting [23], [24]. Other applications (mostly with monochrome

images) such as face recognition [25], [26], [27], [28], character recognition [29] and sign recognition [30] are some of this examples. The proposed approach based on a recursive application of the trace transform to reduce the dimensionality of the obtained feature space, offers an excellent performance for image fingerprinting, but does not offer good discriminative characteristics as a method for domain characterization due to the high data loss occurred in the diametrical and circus functionals [31].

The approach proposed by Liu and Wang [27] reduces the number of attributes using Principal Component Analysis (PCA) to select the most relevant coefficient and reduce the dimensionality of the feature space. However, this approach does not take into account the frequencial relationships among the different coefficients and increases the feature extraction complexity since it requires the covariance matrix information of all previous samples. Moreover, the feature relevance of each individual DCT coefficient is too low and sensitive to noise and variations.

In the following sections a new method for context categorization based on the use of the trace transform will be presented. This method provides higher discriminative characteristics at a very low dimensionality, a key factor for efficient retrieval in massive content databases [32] [33].

This paper is organized as follows: In Section 2 a general overview of our proposed DITEC method is presented. In Section 2.1, image pre-processing issues are addressed. The trace transform and its properties are analyzed in Section 2.2. Feature extraction process details are presented in Section 2.3 while the classification process is described in Section 2.4. The validation carried out with two different datasets is explained in Section 3. Finally, Section 4 concludes with a discussion of the results.

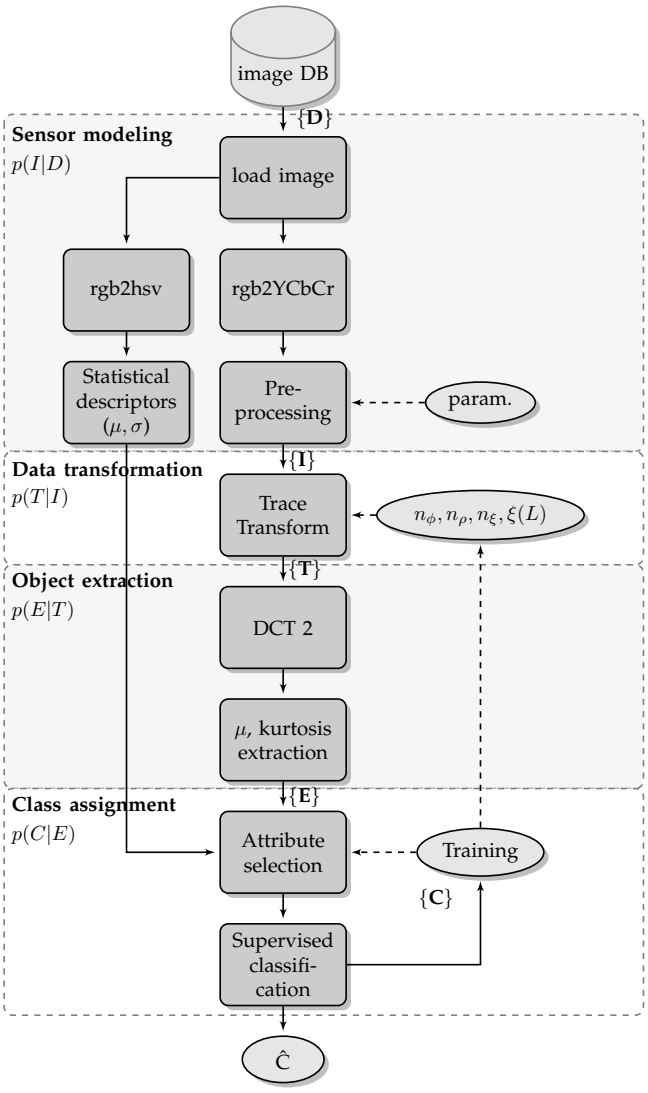


Fig. 1. DITEC System workflow

2 General Description of the DITEC method

We introduce a hierarchical probabilistic model in terms of random variables D, I, T, E and C. The fundamental objective of DITEC is to derive an appropriate estimate $\hat{C}$ of the unknown global image semantic concept C from an observed data set D (Figure 1). Geometric and radio/colorimetric indeterminacies are treated by introducing the concept of an unknown "clean" image I whose parameters depend on the elementary scene descriptors T that depend on scene content E that in turn depends on context C. The conditional probabilistic links between the different layers in the workflow correspond to the main processing steps of the DITEC method.

The four DITEC steps are thus the following:

Sensor modeling: image acquisition and pre-processing (radiometric noise, color space, geometric quantization and image lattice finiteness effects).

Data transformation: Clean image contents in terms of the scene elements in I by means of a trace transform operation with T as outcome of the process. The result will depend on the chosen functional (e.g: (14)) and on the selected geometric parameters (detailed in Section 2.2.3). The outcome of the trace transform of an image is a two-dimensional signal composed of sinusoidal waves. The original image is represented in the resulting signal in terms of sinusoids with a particular amplitude, phase, frequency and intensity. This characterization process represents one of the key steps in the overall information extraction process.

Feature extraction: summarization of the extracted features T, compressed and adapted into a manageable set E of object-based descriptors. The wave features contained in the resulting image must be characterized. In order to do this, the 2D trace signal T_k is transformed to the frequency domain. To concentrate the signal energy to the lowest spatial fre-

quencies, a two-dimensional DCT (Discrete Cosine Transform) is applied. Then, the DCT is compressed to a vector of two components (average value and kurtosis of all the orthogonal elements of the main diagonal, Figure 6). This transformation considers the DCT space as representable by a superposition of Gaussian-shaped clusters. It aims at reducing the considered descriptor space dimensionality while preserving essential information in order to allow a good performance in the subsequent classification process. The last n values from the obtained data pair vector can be disregarded due to the empirical reason that given the low-pass filtering for most natural images the DCT concentrates the highest values in the lowest coefficients.

Class assignment: vectors obtained in the previous step are processed to improve the performance of classifiers in the defined feature space. All the obtained vectors are statistically analyzed to select their most representative attributes. Then the supervised classification process is carried out to obtain an estimate $\hat{C}$ of the unknown global image semantic concept C.

By applying the probability chain decomposition rule, the probability (1) of an asset to belong to a given class can be decomposed in terms of the different layers of the model. Estimates for $C_i, E_j, T_k I_l$, and D_m are the obtained results for $p(C|E)$, $p(E|T)$, $p(T|I)$ and $p(I|D)$ processes, given the usual conditional independence assumptions implied by a hierarchical model:

$$\begin{aligned} &p(C_i|D_m) = p(C_i|E_j, T_k, I_l, D_m) \\ &= p(C_i|E_j)p(E_j|T_k)p(T_k|I_l)p(I_l|D_m)p(D_m) \end{aligned} \tag{1}$$

where:

$$\begin{array}{ccccccc} 0 & < & i & \leqslant & n_{classes} & m \in \mathbb{N} \\ 0 & < & j & \leqslant & \infty & l \in \mathbb{N} \\ 0 & < & k & \leqslant & \infty & k \in \mathbb{N} \\ 0 & < & l & \leqslant & n_{orig.images} & j \in \mathbb{N} \\ 0 & < & m & \leqslant & n_{orig.images} & i \in \mathbb{N} \end{array} \tag{2}$$

$p(C_i|E_j)$ is the probability of the data mining processes to determine correctly the class to which the image belongs, given E as a set of features. The second element $p(E_j|T_k)$ can be understood as $\frac{p(T_k|E_j)p(E_j)}{p(T_k)}$ following the Bayes' theorem. It shows that this model layer is linked to the information representativeness of the extracted features. $p(T_k|I_l)$ implies the trace transform. It is a deterministic process with a slight denoising effect. The quality of data D_m and the pre-processed I_l image will be fundamental for an effective feature extraction process. In fact, the joint inference/estimation process depends on the trace transform which can be regarded as a data re-ordering→compression→feature space optimization process.

2.1 Sensor modeling

The first pre-processing step transforms the RGB color space into YC_bC_r [34]. The luminance channel (Y) will be used as the most relevant channel to encode shape related features. Color distribution information is encoded by processing the chrominance channels (C_b, C_r).

In order to reduce effects introduced by radiometric noise, image lattice and quantization,a low-pass filter is applied to each channel.

HSV [34] color space information is encoded by obtaining mean and variance values (μ, σ) of the corresponding intensity distributions in each H,S,V channel. In the Attribute Selection process, this (μ, σ) information is introduced into the obtained descriptor E.

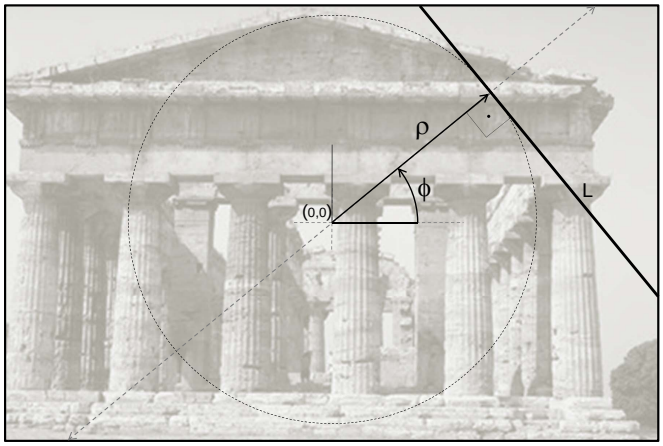


Fig. 2. Trace transform, geometrical representation

2.2 Data transformation

The data transformation process is carried out through the trace transform, a generalization of the Radon transform (3) where the integral of the function is substituted for any other functional Ξ [30], [31], [35], [36], [37].

$$R(\phi, \rho) = \iint f(x, y)\delta(x\cos\phi + y\sin\phi - \rho)dx\ dy \tag{3}$$

The trace transform consists in applying a functional Ξ along a straight line (L in Figure 2). This line is moved tangentially to a circle of radius ρ covering the set of all tangential lines defined by ϕ. The Radon transform has been used to characterize images [38] in well defined domains [39], in image fingerprinting [40] and as a primitive feature for general image description. The trace transform extends the Radon transform by enabling the definition of the functional and thus enhancing the control on the feature space. These features can be set up to show scale, rotation/affine transformation invariance or high discriminance for specific content domains.

The outcome of the trace transform of a 2D image is another 2D signal composed by a set of sinusoidal shapes that vary in amplitude, phase, frequency, intensity and thickness. These sinusoidal signals encode the original image with a given level of distortion depending on the functional and quantization parameters.

2.2.1 Functionals

A functional Ξ of a function $\xi(x)$ evaluated along the line L will have different properties depending on the features of function $\xi(x)$ (e.g.: invariance to rotation, translation and scaling [41]). Kadirov et al. [42] propose several functionals with different invariance or sensitiveness properties. These invariant functionals have been used for expert systems for traffic sign recognition [30], face authentication [26], [43] or fingerprinting [31] purposes.

2.2.2 Geometrical constraints

The main parameter of the trace transform is the functional Ξ while its properties will set the invariant behavior of the transform with respect to its invariance in the face of different image transformations. However, there are geometrical parameters that also have a strong effect on the results. These parameters are the three measures of resolution denoted by $\Delta\phi, \Delta\rho, \xi(\Delta L)$ for angle, radius and along the line L respectively.

The final resolution of the image obtained through the trace transform will be defined by n_ϕ and n_ρ where:

$$n_\phi = \frac{2\pi}{\Delta\phi} \tag{4}$$

$$n_\rho = \frac{\min(X,Y)}{\Delta\rho} \tag{5}$$

with X and Y denoting the horizontal and vertical resolutions of the image I_l.

Low (n_ϕ, n_ρ, n_ξ) values will have a non-linear downsampling effect on the original image, where n_ξ is defined as:

$$n_\xi = \frac{1}{\Delta L} \tag{6}$$

The set of points used to evaluate each functional is described (assuming (0,0) as the center of the image) by:

$$L \rightarrow y = 2\rho\sin(\phi) - \frac{x}{\tan(\phi)} \tag{7}$$

A singularity can be observed at $\phi = 0$ and $\phi = \pi$. For these cases it can be assumed that:

$$L \rightarrow \begin{array}{llll} x = & \rho & \forall y & if \quad \phi = 0 \\ x = & -\rho & \forall y & if \quad \phi = \pi \end{array} \tag{8}$$

The range of the parameters is :

$$\phi \in [0, 2\pi] \tag{9}$$

$$\rho \in [-r, r], r = \min\left(\frac{\frac{X}{2}}{\cos(\phi)}, \frac{\frac{Y}{2}}{\sin(\phi)}\right) \tag{10}$$

$$x \in \left\{ \begin{array}{lll} \left[-\frac{X}{2}, \frac{X}{2}\right] & \forall & \phi \in \alpha \\ \left[-\frac{\frac{Y}{2}}{\tan(\phi)}, \frac{\frac{Y}{2}}{\tan(\phi)}\right] & \forall & \phi \in \beta \end{array} \right\} \tag{11}$$

$$\alpha \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right] \cup \left[-\frac{3\pi}{4}, \frac{5\pi}{4}\right]$$

$$\beta \in \left(\frac{\pi}{4}, \frac{3\pi}{4}\right) \cup \left(\frac{5\pi}{4}\frac{7\pi}{4}\right)$$

X and Y are the horizontal and vertical resolutions of the image. Equation (7) shows a symmetrical result since the same lines are obtained for $\phi \in [0, \pi]$ and $\phi \in [\pi, 2\pi]$. However this is only true for functionals that are not considering the position (like the Radon transform). Depending on the selected functional and on the desired properties of the trace transform(e.g: rotational invariance), the ranges of ϕ and ρ can be modified to: $\phi \in [0, \pi]$ or $\rho \in [0, r]$.

2.2.3 Quantization effects

Digital images are affected by two main effects during trace transformation:

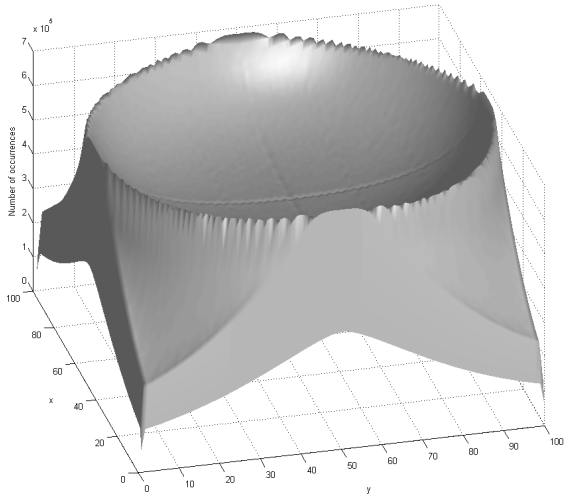


Fig. 3. Trace transform contribution mask at very high resolution parameters (Image resolution:100x100px. $n_\phi = 1000, n\rho = 1000, n_\xi = 5000$).

- some pixels might never be used by the functional given the geometrical setup of the transform, and to its integration nature.
- there may be some pixels that have much higher cumulated effect than the others into the functional.

These effects need to be taken into account in order to preserve of the results the homogeneity, avoiding pixels or areas with higher relevance than others. Even for very high (n_ϕ, n_ρ, n_ξ) values in relation to the original image

resolution, the trace transform introduces a contribution intensity map that encodes the relevance of the different regions of the input picture. As shown in Figure 3, high resolution values of the trace transform parameters tend to create a convex contribution intensity map. Therefore, high parameter values do not necessarily imply optimal image content representation on the trace transform.

High values of n_ϕ improve the rotational invariance of the trace transform (although in a manner that it is dependent on the selected functional) while very low values $n_\phi < 5$ cannot be considered as producing a valid trace transform since there is not enough angular information.

Ideally, the trace transform should keep the following constraints (considering M as the matrix that contains the number of repetitions of each pixel during the trace transform):

- **Coverage** All pixels of the image have to be included at least in one functional. $\min(M) > 0$.
- **Homogeneity** All pixels are used the same number of times. Var$(M) = 0$.
- **High pixel repetition degree** Each pixel has to be included in as many traces as possible (high values of $mean(M)$).

TABLE 1
Quantization effects of the trace transform

n_ϕ	n_ρ	$n_{F(L)}$	% pixels used	Mean	Var
64	64	15	16.60	0.63	15.71
64	64	45	44.30	1.88	32.72
64	64	85	67.53	3.54	53.61
64	64	185	93.40	7.71	52.51
300	5	45	28.62	0.69	10.28
300	5	151	69.84	2.30	31.80
5	300	45	40.59	0.68	0.20
5	300	151	88.43	2.30	0.42
5	300	218	97.34	3.33	0.40
5	300	251	99.18	3.83	0.30
384	256	15	83.76	15.00	$1.2{\cdot}10^6$
100	100	85	85.55	8.65	872.47
100	100	185	98.72	18.82	708.64
100	100	218	99.55	22.18	511.61
100	100	2,185	100.00	222.27	$3.6{\cdot}10^6$
42	75	12,000	99.77	384.52	$38.6{\cdot}10^6$

Table 1 shows some example values for coverage, homogeneity and repetition degree at different n_ϕ, n_ρ, n_ξ resolutions. Note that the best ratios are obtained for lower variations in ϕ since the angle is the main factor to increase the variance. The pixel repetition degree is also strongly conditioned by the angular resolution. This fact makes n_ϕ the main factor to balance the homogeneity and repetition degree (e.g: low repetition degrees show weaker rotational invariance). Once n_ϕ is set, n_ρ can be adjusted to ensure the optimal coverage. n_ξ has an almost asymptotic behavior once the other two parameters are set. Figure 4 shows some cases applied to a real image and the convex contribution intensity mask effect for moderate or higher values of n_ϕ.

2.3 Feature extraction

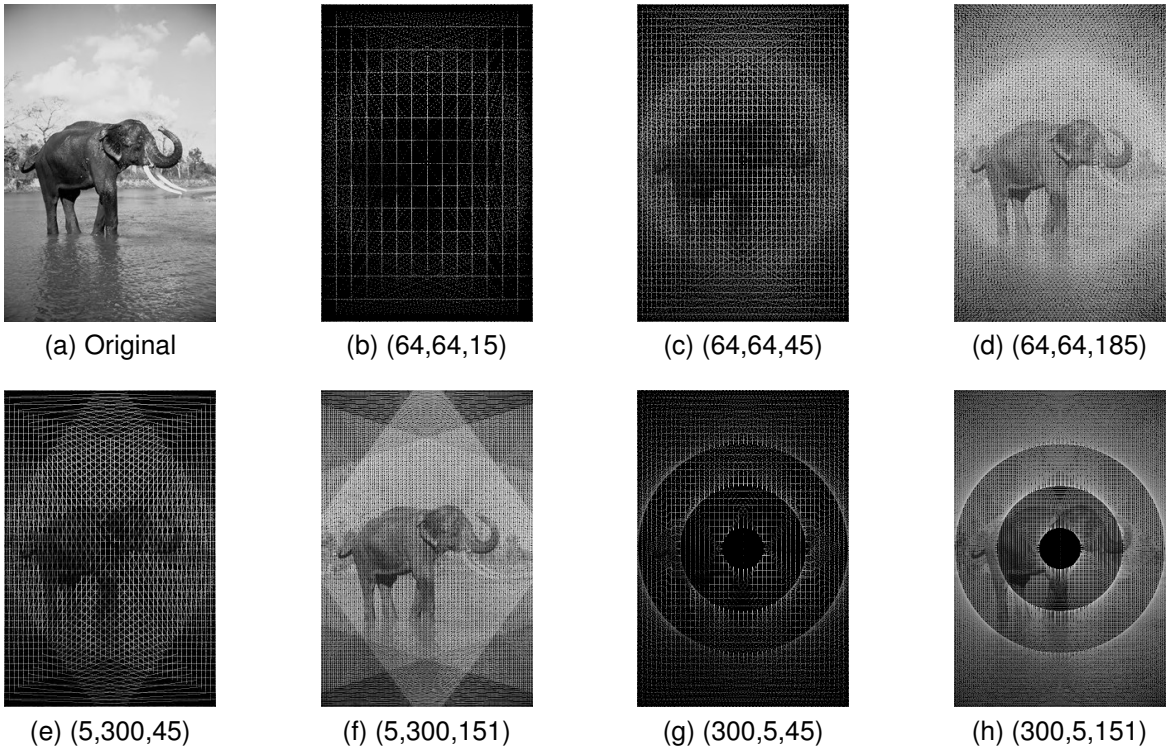
In order to reduce the set of descriptors that are needed to characterize the wave-like signal obtained from the trace transform, a DFT Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) can be applied. The DCT [44], which has become one of the most popular transforms for audio and image coding, has two main properties that make it more suitable than DFT for the feature extraction process: *energy compaction* and *decorrelation* [45]. The energy compaction means that the signal energy is accumulated in a small number of coefficients and that these coefficients are typically the lowest coefficients of the DCT transform. Taking into account that the trace transform does not introduce high frequencies into the transformed image, the DCT provides a good method to efficiently represent the wave-like signal information contained in the resulting images. The decorrelation property of the DCT implies that there is a very low interdependency among the coefficients. This property matches with the common needs of a number of data mining algorithms whose performance has a strong dependency on input attribute correlation. Moreover, the coefficients obtained by applying a DCT are real values while the DFT provides coefficients in the complex domain. The DCT thus allows to encode information in lower dimensionality code spaces with better compaction characteristics. Moreover, from the computational cost point of view, there are efficient SW and HW algorithms for the implementation of the DCT that make it suitable for real time applications without high computing performance requirements.

The 2D forward DCT is given by (12). In our case, instead of using the typical 8x8 macroblocks (which improve the coding speed but act as local descriptors), the transform will be applied to the whole image, keeping the global representativeness of the obtained coefficients.

$$X_{k_1k_2} = \alpha_{k_1}\alpha_{k_2}\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x_{n_1n_2}\cos\left[\frac{\pi k_1(2n_1+1)}{2N_1}\right]\cdot \cos\left[\frac{\pi k_2(2n_2+1)}{2N_2}\right] \tag{12}$$

where:

$$\alpha \in \begin{cases} \frac{1}{\sqrt{N_i}} & k_i = 0 \\ \sqrt{\frac{2}{N_i}} & k_i \neq 0 \end{cases} \tag{13}$$



(a) Original (b) (64,64,15) (c) (64,64,45) (d) (64,64,185)

(e) (5,300,45) (f) (5,300,151) (g) (300,5,45) (h) (300,5,151)

Fig. 4. Pixels relevance in trace transform scanning process with different parameters (n_ϕ, n_ρ, n_ξ). Original image resolution = 384x256.

2.3.1 Statistical descriptors

As a consequence of the properties of the DCT and of the nature of the 2D signals resulting from the trace transform, the 2D DCT stores more energy in its lower frequencies.

Figure 5 shows the process of trace transform evaluation and its 2D DCT where the intensity is quantized into 5 different levels. The functional used has been the one enumerated by Srisuk et al. [26] as functional number 3 (14).

$$T(f(t)) = \int_c^\infty (t-c)^2 f(t) dt \tag{14}$$

$$c = \frac{1}{S} \int_0^\infty t \left| f(t) \right| dt \tag{15}$$

$$S = \int_0^\infty \left| f(t) \right| dt \tag{16}$$

In order to avoid this sensitivity to specific coefficients of the DCT and instead of the previously discussed PCA based approach, our proposed DITEC method for dimensionality reduction is based on statistical parameters of the n first perpendicular straight lines to the main diagonal (Figure 6). These coefficients which correspond to similar frequency bands can be computed very efficiently. The distribution is represented by the mean value and the kurtosis of each vector. This pair of descriptors (μ, k) of the first element (corresponding to the DC value of the DCT) is substituted by the mean and variance of the original image in HSV space (Figure 1).

Equation (17) defines the kurtosis of a distribution which is represented by (18) for a discrete set of elements.

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{17}$$

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \tag{18}$$

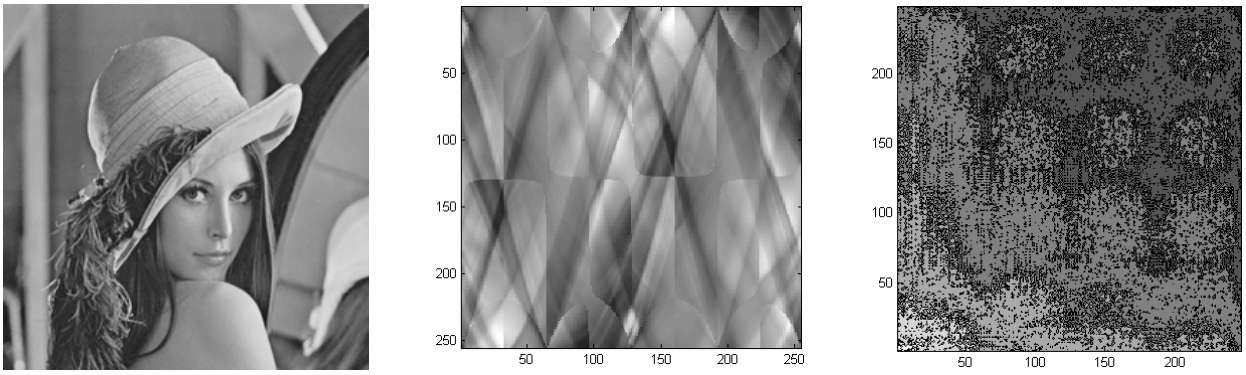
Consider that the mean and kurtosis values encode the information of coefficients corresponding to approximately similar frequencies. The obtained dimensionality of the transformed (μ, k) pairs is given by (19).

$$nDims = \sqrt{n_\phi^2 + n_\rho^2} \cdot n_c \cdot n_f \tag{19}$$

where n_c is the number of channels of the original image and n_f the number of features extracted from each vector (2 in the case of using $[\mu, k]$). Thus, the dimensionality reduction is given by (20).

$$rf = \frac{n_\phi n_\rho}{\sqrt{n_\phi^2 + n_\rho^2} \cdot n_f} \tag{20}$$

For square resolutions and considering $n_f = 2$ the reduction factor increases linearly with the resolution (21).



(a) Original (b) Trace transform (F3) (c) DCT2 4 level quantization

Fig. 5. Trace Transform and subsequent Discrete Cosine Transform of Lenna. (Y channel of YCbCr color space)

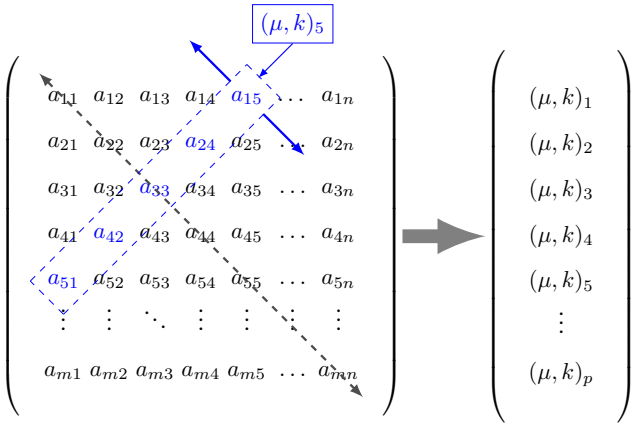


Fig. 6. Conceptual scheme: DCT matrix transformation into μ, k pair vector.

$$rf = \frac{n^2}{n \cdot n_c\sqrt{2}} = \frac{n}{2\sqrt{2}} \tag{21}$$

2.4 Classification

After the feature extraction process explained in the previous section, the dimensionality of resulting descriptors can be reduced by attribute selection strategies in order to improve the efficiency of subsequent classification steps.

Considering machine learning as a set of techniques to discover and extract knowledge in an automated way [46], the basic problem is concerned with the induction of a model that classifies a given object into one of several known classes. In order to induce the classification model, each element E described by a pattern of d features is simplified by applying the Feature Subset Selection (FSS) [47] approach. FSS can be reformulated as follows: *given a set of candidate features, select the "best" subset in a classification problem*. In our case, the "best" subset will be the one with the best predictive accuracy.

Most of the supervised learning algorithms perform rather poorly when faced with many irrelevant or redundant (depending on the specific characteristics of the classifier) features. In this way, the FSS proposes additional methods to reduce the number of features so as to improve the performance of the supervised classification algorithm.

2.4.1 Feature Subset Selection in Machine Learning

There are two main approaches to tackle the Feature Subset Selection (FSS) problem from the Machine Learning point of view [48], namely wrapper and filter methods.

Wrapper approaches [49] try to identify the subset of variables that, given a classification paradigm and a dataset, provide the best classification function. The process consists on searching an optimal feature subspace based on a performance measure (typically the accuracy, though other measures can be used). Each subset is evaluated by testing the performance of the chosen paradigm in the dataset, using only the variables in the subset for evaluation. The estimation of the performance of the classifiers requires a validation scheme, such as cross validation or bootstrap estimation. As a result, the evaluation of each subset involves the training and testing of several classification functions, increasing the computational time required for the FSS process.

The filter approaches search for the best variable subset, independently of the classification paradigm, considering the relationship between the predicting variables and the class, and occasionally the relationship among the predicting variables. One of the simplest approaches consists of ranking the variables according to their usefulness and selecting only those on the top of the ranking. The usefulness of a variable is measured univariately by means of different metrics.

Once the features are ranked, a threshold must be set to obtain the final subset. The ranking methods are only concerned with the relevance of the features considered and, thus, they do not filter out redundant variables.

The selected classifiers are briefly described below; a wrapper Feature Subset Selection has been used in this paper.

For the supervised learning task, in the training set used to generate the classification model, for each x

sample its y label value is known. For this analysis, Bayesian Networks [50] and Support Vector Machines (SVM) [51] have been used.

3 Experimental Results

The presented method has been tested with 2 different datasets. The first of them (Corel 1000 [52]) is a standard dataset which will allow the comparison of the obtained validation data with other methods existing in the literature. The second case (earth observation data), will be used to show the potential of the proposed method under diverse conditions.An *a priori* statistical data analysis together with a combination of classifiers has been adapted for each of the two corresponding validation case studies.

3.1 Case study 1: Corel 1000 dataset

The Corel 1000 dataset is composed of 1000 images distributed in 10 classes (100 instances per class). The tags of the classes are: *Africans, Beach, Architecture, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food*. Figure 7 shows one sample per each class. Even though they are semantically separated, visual similarities may be found among some of them. For example, people and trees can be found under *Africa, Beach,* and *Mountain* categories.

The following parameters have been selected: $n_\phi = 71, n_\rho = 71, n_c = 3, n_f = 2$. This choice results in 15,123 trace transform coefficients per image. By obtaining the mean values and kurtosis as described in the previous section, the number of attributes is reduced to 606 (by a factor of 25).

Based on the fact that the DCT gathers signal energy in the lower frequencies (see Figure 5c), highest coefficients are removed. Moreover, it can be assumed that chrominance channels (C_b and C_r) contain less visual information and therefore more coefficients can be removed from these channels than from the luminance signal (Y). Experimental results carried out with different combination of YC_bC_r coefficients, demonstrate that luminance related attributes have more relevance than chrominance related ones. The selected parameters for this example result in 202 attributes per channel. We will select the first 104 ones for Y and 60 for each C_bC_r signal, thus reducing the total amount of attributes to 224.

The best performance has been obtained by applying a SVM classifier (precision = 84.8% in a k-fold 10 test). 117 attributes have been selected for the final feature space. The information provided by the confusion matrix (Table 2) can be represented graphically in order to represent the qualitative behaviour of the method. We have selected the *Force Atlas 2* algorithm [53] to distribute the classes on a 2D plane. *Force Atlas 2* establishes a force directed layout simulating a physical system where nodes (classes) repulse each other and edges apply an attraction force. For the method presented in this paper, the repulsion force is adjusted to scale the layout to a convenient size while edge forces are represented by the error information stored in the confusion matrix. Thus, the attraction force of two nodes will be proportional to the mutual miss-classifications.

TABLE 2
Corel 1000 dataset confusion matrix

	a	b	c	d	e	f	g	h	i	j
a	75	2	6	0	2	5	0	2	1	7
b	5	79	6	1	0	6	0	0	2	1
c	3	4	78	1	0	3	1	0	8	2
d	3	3	3	81	0	0	1	0	4	5
e	0	0	0	0	100	0	0	0	0	0
f	7	1	3	0	0	83	0	2	3	1
g	1	1	0	0	0	0	95	2	0	1
h	1	0	1	1	0	0	0	97	0	0
i	0	14	4	1	0	3	0	0	78	0
j	5	1	0	5	0	3	4	0	0	82

For the Corel 1000 dataset, it can be observed in Figure 8 that *Dinosaurs, Flowers* and *Horses* are clearly separated from the rest of the categories. This result can also be verified via the precision and recall data. Precision is above 95% and there are very few instances for other classes estimated as *Dinosaurs, Flowers* or *Horses*.

A deeper analysis of class distribution can be performed by removing the aforementioned three categories. Figure 9 shows that there is a group formed by *Beach, Mountains* and *Architecture* and other by *Africans* which links to *Elephants* and *Food* although these two are not directly connected.

Figure 10 shows an example of one of the classification errors. As it can be seen, the presence of vegetation and trees associates the image to the *Mountain* class even if it belongs to *Architecture*. These semantic overlays of Corel 1000 categories put some visually similar images in different classes.

Comparing the obtained results with other feature extraction approaches (Mean-Shift and Gaussian Mixtures based on Weighted Color Histograms [12], Reduced Feature Vector with Relevance Feedback [54] and SIFT based Gaussian Naïve Bayesian Network [55]), DITEC shows the best performance for most categories (Figure 11) and the highest mean precision value.

3.2 Case study 2: Geoeye satellite imagery

The Geoeye [56] dataset is composed by 1003 multiresolution patches of Digital Globe Earth observation satellite imagery at $\sim$ 1m spatial resolution. The dataset is categorized in 7 classes corresponding to different geographical locations (Figure 12). All the resolutions have been processed with the same trace transform parameters.

During the data mining process Bayesian networks have shown the best performance, reaching a precision of94.51% in a k-fold 10 test. The final dimensionality of

(a) Africans (b) Beach (c) Architecture (d) Buses (e) Dinosaurs

(f) Elephants (g) Flowers (h) Horses (i) Mountains (j) Food

Fig. 7. Samples of Corel 1000 dataset. The dataset includes 256x384 or 384x256 images.

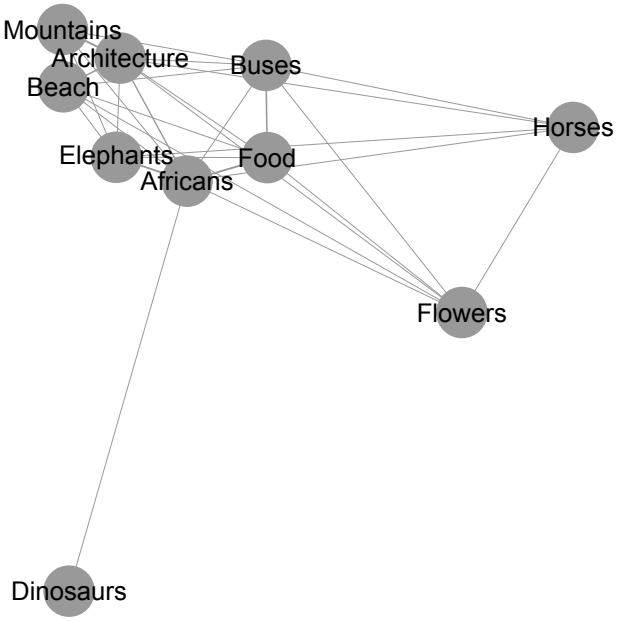


Fig. 8. Distance among classes in the Corel 1000 dataset according to misclassified instances.

the feature space has been reduced to 61 attributes. Table 3 shows the confusion matrix of the classification results.

TABLE 3
Geoeye dataset confusion matrix

	a	b	c	d	e	f	g
(a) Athens	74	0	1	0	2	0	0
(b) Davis	0	183	0	0	2	7	2
(c) Manama	1	0	193	0	0	0	0
(d) Midway	2	0	0	62	1	0	0
(e) Nyragongo	0	0	4	0	77	2	2
(f) Risalpur	0	0	0	0	0	177	17
(g) Rome	0	0	1	0	0	11	182

Applying the *Force Atlas 2* method to Geoeye classification errors, we obtain the distribution shown in

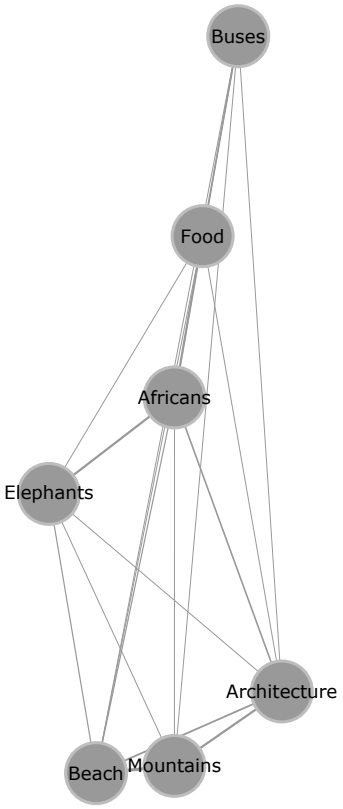


Fig. 9. Distance among most inter-related classes in the Corel 1000 dataset according to misclassified instances.

Figure 13. It can be observed that *Risalpur* and *Rome* are the categories with the highest mutual similarity (2 cities). The Davis-Monthan aircraft boneyard has shown a remarkable similarity with Risalpur due to the fact that wide areas of bare soil are a common element in both *Risalpur* and *Davis*.

The Midway atoll is the most distinguishable category of the Geoeye dataset. It has special color features and textures and shapes are also singular within the dataset. All these characteristics have been successfully detected

Fig. 10. Corel 1000 picture corresponding to class *Architecture* and classified as *Mountain*

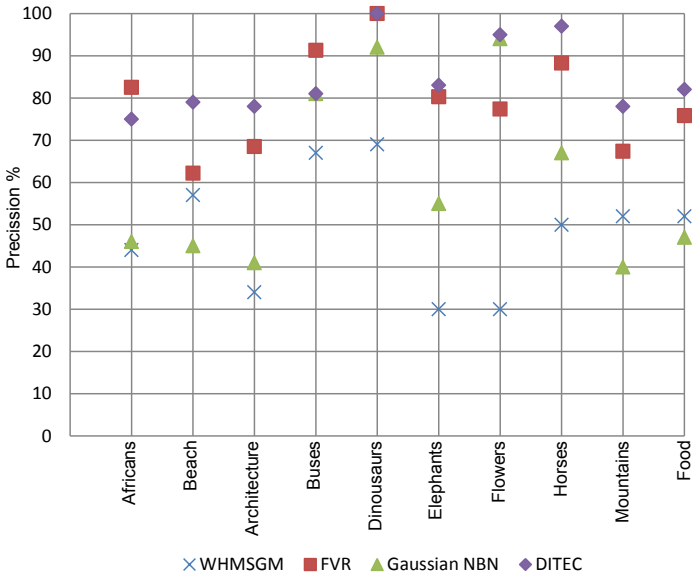


Fig. 11. Corel 1000 precision results with different feature extraction algorithms. *WHMSGM*: Mean-Shift and Gaussian Mixtures based on Weighted Color Histograms, *FVR*: Reduced Feature Vector with Relevance Feedback, *Gaussian NBN*: SIFT based Gaussian Naïve Bayesian Network.

by the method (Precision = 100%, Recall = 0.954%).

4 Conclusion

We have shown that the trace transform provides highly discriminant features for context categorization purposes that can be encoded as considerably short feature vectors. We have presented the geometrical constraints of the trace transform that can be optimized to efficiently represent the information contained in the original images. The dimensionality reduction in terms of mean and kurtosis value pair of frequencial coefficients results in a very robust set of features in terms of precision. For most resolution $(n_\phi, n_\rho, L(n))$ settings maintaining acceptable coverage, homogeneity and redundancy conditions, precision has demonstrated to keep around 82% for the Corel 1000 dataset and 92% for Geoeye.

Moreover, the method has successfully identified visual similarities within the datasets, and as seen in the validation section, some incorrectly classified instances are in fact visually similar to those pointed out by the classifier. The error analysis has also shown some semantic proximity between visually similar categories, a fact that can be used for context modeling and automatic ontology building.

References

[1] A. Torralba, K. P. Murphy, and W. T. Freeman, "Using the forest to see the trees: exploiting context for visual object detection and localization," *Commun. ACM*, vol. 53, no. 3, pp. 107–114, Mar. 2010. [Online]. Available: http://doi.acm.org/10.1145/1666420.1666446

[2] A. Torralba and P. Sinha, "Statistical context priming for object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2001. [Online]. Available: http://web.mit.edu/torralba/www/iccv2001.pdf

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000. [Online]. Available: http://dx.doi.org/10.1109/34.895972

[4] G. Marcos, K. Alonso, I. G. Olaizola, J. Florez, and A. Illarramendi, "Dms-1 driven data model to enable a semantic middleware for multimedia information retrieval in a broadcaster," in *Proc. 4th Int. Workshop Semantic Media Adaptation and Personalization SMAP '09*, 2009, pp. 33–37.

[5] C. G. M. Snoek and A. W. M. Smeulders, "Visual-concept search solved?" *Computer*, vol. 43, no. 6, pp. 76–78, 2010.

[6] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE_J_PAMI*, vol. 32, no. 7, pp. 1271–1283, 2010.

[7] D. G. Lowe, "Object recognition from local Scale-Invariant features," *Computer Vision, IEEE International Conference on*, vol. 2, pp. 1150–1157 vol.2, Aug. 1999. [Online]. Available: http://dx.doi.org/10.1109/ICCV.1999.790410

[8] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *In ECCV*, 2006, pp. 404–417.

[9] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *Proc. IEEE 11th Int. Conf. Computer Vision ICCV 2007*, 2007, pp. 1–8.

[10] M. J. Choi, J. J. Lim, A. Torralba, and A. S. Willsky, "Exploiting hierarchical context on a large database of object categories," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 129–136.

[11] I. G. Olaizola, G. Marcos, P. Kramer, J. Florez, and B. Sierra, "Architecture for semi-automatic multimedia analysis by hypothesis reinforcement," in *Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting BMSB '09*, 2009, pp. 1–6.

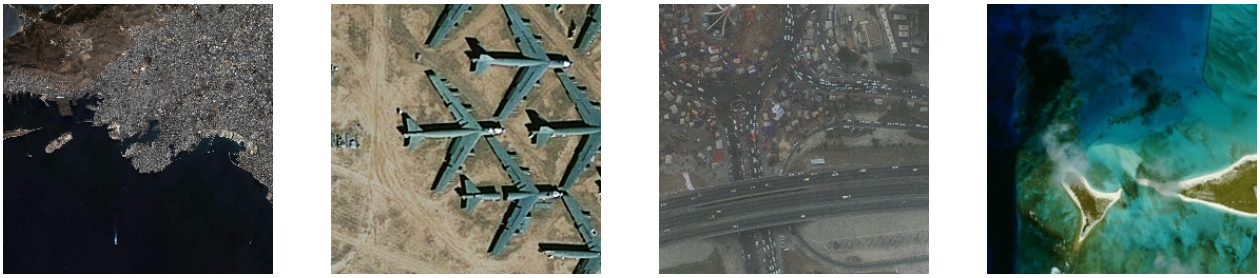
[12] M. A. Bouker and E. Hervet, "Retrieval of images using mean-shift and gaussian mixtures based on weighted color histograms," in *Proc. Seventh Int Signal-Image Technology and Internet-Based Systems (SITIS) Conf*, 2011, pp. 218–222.

[13] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR '07*, 2007, pp. 1–8.

[14] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, pp. 145–175, 2001.

[15] A. Torralba, R. Fergus, and Y. Weiss, "Small codes and large image databases for recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR 2008*, 2008, pp. 1–8.

[16] T. Watanabe, K. Sugawara, and H. Sugihara, "A new pattern representation scheme using data compression," *IEEE_J_PAMI*, vol. 24, no. 5, pp. 579–590, 2002.



(a) Athens (b) Davis (c) Manama (d) Midway

(e) Nyragongo (f) Risalpur (g) Rome

Fig. 12. Samples of satellite footage dataset. 256x256px patches at different scales.

[17] S. Ahmed, M. Khan, and M. Shahjahan, "A filter based feature selection approach using lempel ziv complexity," in *Advances in Neural Networks ISNN 2011*, ser. Lecture Notes in Computer Science, D. Liu, H. Zhang, M. Polycarpou, C. Alippi, and H. He, Eds. Springer Berlin / Heidelberg, 2011, vol. 6676, pp. 260–269, 10.1007/978-3-642-21090-7_31. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21090-7_31

[18] D. Cerra, A. Mallet, L. Gueguen, and M. Datcu, "Algorithmic information theory-based analysis of earth observation images: An assessment," *IEEE_J_GRSL*, vol. 7, no. 1, pp. 8–12, 2010.

[19] M. R. Mustaffa, F. Ahmad, R. Mahmod, and S. Doraisamy, "Generalized ridgelet-fourier for mxn images: Determining the normalization criteria," in *Proc. IEEE Int Signal and Image Processing Applications (ICSIPA) Conf*, 2009, pp. 380–384.

[20] ——, "Invariant generalised ridgelet-fourier for shape-based image retrieval," in *Proc. Int Information Retrieval & Knowledge Management, (CAMP) Conf*, 2010, pp. 79–84.

[21] H. Nemmour and Y. Chibani, "Handwritten arabic word recognition based on ridgelet transform and support vector machines," in *High Performance Computing and Simulation (HPCS), 2011 International Conference on*, july 2011, pp. 357 –361.

[22] *MPEG-7 Overview*, ISO/IEC JTC1/SC29/WG11 Std., Rev. 10, October 2004. [Online]. Available: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

[23] M. Bober and R. Oami, "Description of mpeg-7 visual core experiments," ISO/IEC JTC1/SC29/WG11, Tech. Rep., 2007.

[24] R. O'Callaghan, M. Bober, and P. Oami, R. and. Brasnett, *Information technology - Multimedia content description interface - Part 3: Visual, AMENDMENT 3: Image signature tools*, ISO/IEC Std. ISO/IEC TC JTC1/SC 29/WG 11 N 9581, 01 2008.

[25] S. A. Fahmy, "Investigating trace transform architectures for face authentication," in *Proc. Int. Conf. Field Programmable Logic and Applications FPL '06*, 2006, pp. 1–2.

[26] S. Srisuk, M. Petrou, W. Kurutach, and A. Kadyrov, "Face authentication using the trace transform," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 1, 2003.

[27] N. Liu and H. Wang, "Recognition of human faces using siscrete cosine transform filtered trace features," in *Proc. 6th Int Information, Communications & Signal Processing Conf*, 2007, pp. 1–5.

[28] ——, "Modeling images with multiple trace transforms for pattern analysis," *IEEE J SPL*, vol. 16, no. 5, pp. 394–397, 2009.

[29] M. F. Nasrudin, M. Petrou, and L. Kotoulas, "Jawi character recognition using the trace transform," in *Proc. Seventh Int Computer Graphics, Imaging and Visualization (CGIV) Conf*, 2010, pp. 151–156.

[30] J. Turan, Z. Bojkovic, P. Filo, and L. Ovsenik, "Invariant image recognition experiment with trace transform," in *Proc. 7th Int Telecommunications in Modern Satellite, Cable and Broadcasting Services Conf*, vol. 1, 2005, pp. 189–192.

[31] A. Kadyrov and M. Petrou, "The trace transform and its applications," *IEEE J PAMI*, vol. 23, no. 8, pp. 811–828, 2001.

[32] P. J. Haas, "Sketches get sketchier," *Commun. ACM*, vol. 54, pp. 100–100, August 2011. [Online]. Available: http://doi.acm.org/10.1145/1978542.1978565

[33] P. Li and C. König, "b-bit minwise hashing," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 671–680. [Online]. Available: http://doi.acm.org/10.1145/1772690.1772759

[34] C. Poynton, *A technical introduction to digital video*. J. Wiley, 1996. [Online]. Available: http://books.google.es/books?id=jI5jQgAACAAJ

[35] A. Kadyrov and M. Petrou, "The trace transform as a tool to invariant feature construction," in *Proc. Fourteenth Int Pattern Recognition Conf*, vol. 2, 1998, pp. 1037–1039.

[36] M. Petrou and A. Kadyrov, "Affine invariant features from the trace transform," *IEEE_J_PAMI*, vol. 26, no. 1, pp. 30–44, 2004.

[37] P. Brasnett and M. Bober, "Fast and robust image identification," in *Proc. 19th Int. Conf. Pattern Recognition ICPR 2008*, 2008, pp. 1–5.

[38] F. Peyrin and R. Goutte, "Image invariant via the radon transform," in *Proc. Int Image Processing and its Applications Conf*, 1992, pp. 458–461.

[39] S. Lin, S. Li, and C. Li, "A fast electronic components orientation and identify method via radon transform," in *Proc. IEEE Int Systems Man and Cybernetics (SMC) Conf*, 2010, pp. 3902–3908.

[40] J. S. Seo, J. Haitsma, T. Kalker, and C. D. Yoo, "A robust image fingerprinting system using the radon transform," *Sig. Proc.: Image Comm.*, vol. 19, no. 4, pp. 325–339, 2004.

[41] R. Fooprateepsiri, W. Kurutach, and S. Tamsumpaolerd, "An image identifier based on hausdorff shape trace transform," in *Proceedings of the 16th International Conference on Neural Information Processing: Part I*, ser. ICONIP '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 788–797. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-10677-4_90

[42] A. Kadyrov and M. Petrou, "Affine parameter estimation from the trace transform," *IEEE J PAMI*, vol. 28, no. 10, pp. 1631–1645, 2006.

[43] Z. Shi, M. Du, and R. Huang, "A trace transform based on subspace method for face recognition," in *Proc. Int Computer Application and System Modeling (ICCASM) Conf*, vol. 13, 2010.

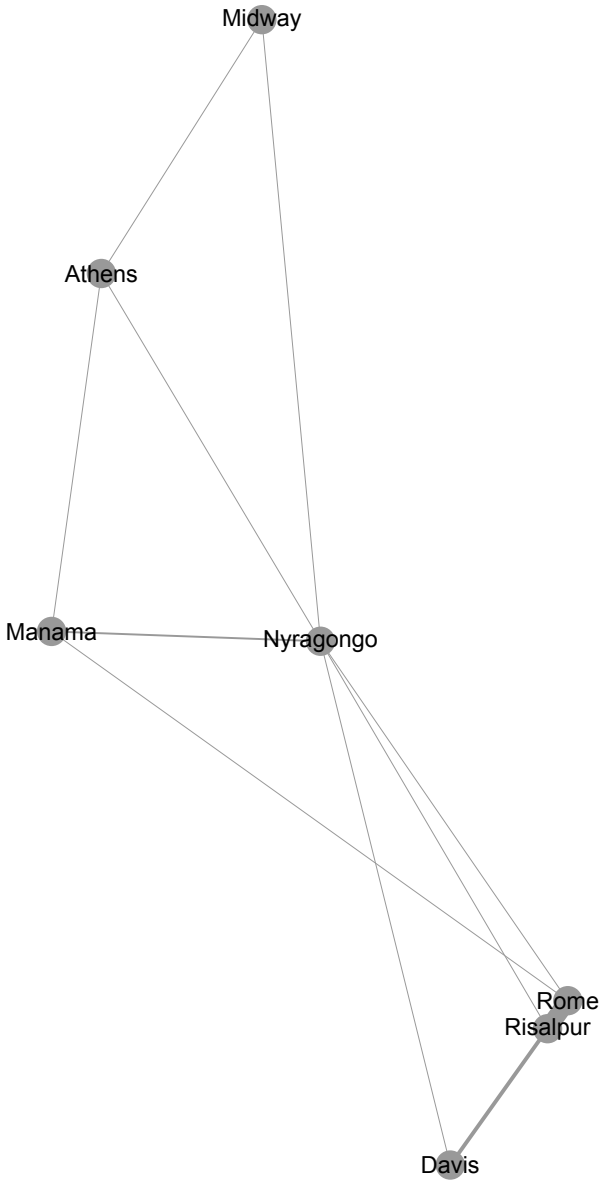


Fig. 13. Distance among classes in the Geoeye dataset according to misclassified instances.

[44] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE-J-C*, no. 1, pp. 90–93, 1974.
[45] I. E. Richardson, *Video Codec Design: Developing Image and Video Compression Systems*. New York, NY, USA: John Wiley & Sons, Inc., 2002.
[46] T. Mitchell, *Machine Learning*. McGraw Hill, 1997.
[47] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, 1998.
[48] I. Inza, P. Larrañaga, and B. S. R. Etxeberria, "Feature subset selection by Bayesian networks based optimization," *Artificial Intelligence*, vol. 123, no. 1–2, pp. 157–184, 2000.
[49] R. Blanco, P. Larrañaga, I. Inza, and B. Sierra, "Gene selection for cancer classification using wrapper approaches," *International Journal of Pattern Recognition and Artificial Intelligence*, 2004.
[50] B. Sierra, E. Lazkano, E. Jauregi, and I. Irigoien, "Histogram distance-based bayesian network structure learning: A supervised classification specific approach," *Decision Support Systems*, vol. 48, no. 1, pp. 180–190, 2009.
[51] D. Meyer, F. Leisch, , and K. Hortnik, "The support vector machine under test," *Neurocomputing*, vol. 55, pp. 169–186, 2003.
[52] [Online]. Available: http://wang.ist.psu.edu/docs/related.shtml
[53] M. Jacomy, S. Heymann, T. Venturini, and M. Bastian, "Forceatlas2, a graph layout algorithm for handy network visualization," Gephi – Web Atlas, Draft, 2011. [Online]. Available: http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf
[54] G. Zajic, N. Kojic, N. Reljin, and B. Reljin, "Experiment with reduced feature vector in cbir system with relevance feedback," *IET Conference Publications*, vol. 2008, no. CP543, pp. 176–181, 2008. [Online]. Available: http://link.aip.org/link/abstract/IEECPS/v2008/iCP543/p176/s1
[55] W. Bouachir, M. Kardouchi, and N. Belacel, "Fuzzy indexing for bag of features scene categorization," in *Proc. 5th Int I/V Communications and Mobile Network (ISVC) Symp*, 2010, pp. 1–4.
[56] [Online]. Available: http://www.geoeye.com

Igor García Olaizola is the head of Digital TV and Multimedia Services Department in Vicomtech (http://www.vicomtech.org). He received his MEng degree in Electronic and Control Engineering from the University of Navarra, Spain (2001). He developed his Master thesis at Fraunhofer Institut für Integrierte Schaltungen (IIS), Erlangen -Germany- 2001 and currently he is preparing his PhD in Computing Science and Artificial Intelligence at University of Basque Country. He has participated in many industrial projects related with Digital TV as well as several European research projects in the area of audiovisual content management. His current research interests include multimedia content analysis frameworks and techniques to decrease the semantic gap.

Marco Quartulli Marco Quartulli Marco Quartulli received the laurea degree in physics from the University of Bari, Italy, in 1997 and a PhD in EE and CS from the University of Siegen, Germany, in 2005. He worked from 1997 to 2010 on remote sensing ground segment engineering, image analysis, archives and mining for Advanced Computer Systems, Italy. From 2000 to 2003, he was with the Image Analysis Group at the Remote Sensing Technology Institute of the German Aerospace Center (DLR) in Germany. Since 2010, he has joined the Digital Television department of Vicomtech in Spain. His research interests include multimedia mining and asset management, multiple acquisition high-resolution optical and radar remote sensing, scene understanding, data fusion and Bayesian modeling.

Julián Flórez Dr. Julián Flórez studied Industrial Engineering in the University of Navarra (1980), and obtained his Ph.D. in the University of Manchester, Institute of Science and Technology UMIST (1985), in the field of Adaptive Control. From 1985 to 1990, he worked as Researcher in the Centre of Study and Technical Research of Gipuzkoa (CEIT), where he collaborated in several research projects related to Electrical and Industrial Engineering with a marked industrial focus. From 1985 to 1994, he was Associate Professor in the School of Industrial Engineering of the University of Navarra, and since 1994, he is Professor at the same university. From 1990 to 1997, Dr. Flórez worked as Senior Researcher in CEIT, where he was in charge of a Department of Industrial Applications. From 1997 to 2001, he worked as Director of Corporate Development of Avanzit-SGT (Servicios Generales de Teledifusión) in the fields of Information Systems, Communications and Broadcasting infrastructure. He has a strong background in Digital Television infrastructures and was tightly involved in the deployment of one of the biggest Digital TV organizations in Spain and Europe: Quiero TV. Since 2001, he is Principal Researcher in Vicomtech. He holds some patents and has written more than 40 research papers in different areas of Industrial and Electrical Engineering.

Basilio Sierra is Full Professor in the Computer Sciences and Artificial Intelligence Department at the University of the Basque Country. He received his BSc in Computer Sciences in 1990, MSc in Computer Science and Architecture in 1992 and PhD in Computer Sciences in 2000 at the University of the Basque Country. He is the director of the Robotics and Autonomous Systems Group in Donostia-San Sebastian. Professor Sierra is presently a researcher in the fields of Robotics, Computer Vision and Machine Learning, and he is working on the development of different paradigms to improve classification behaviors; he has written more than 25 journal papers in those fields, as well as more than 100 conference contributions and more than 30 book chapters.